



## Development of a biomarker scoring system for use in graft-versus-host disease

“This scoring system consistently classified patients into three categories of risk that predicted response to treatment within 28 days as well as risk of mortality from acute graft-versus-host disease...”

First draft submitted: 7 June 2016; Accepted for publication: 14 June 2016; Published online: 14 July 2016

**Keywords:** algorithm • Ann Arbor score • biomarkers • graft-versus-host disease (GVHD) • treatment

The ability of allogeneic hematopoietic stem cell transplantation (HCT) to cure hematological malignancies is due, in large part, to the graft-versus-leukemia (GVL) effect mediated by alloreactive T cells in the donor graft. But GVL effects remain closely associated with graft-versus-host disease (GVHD), which is mediated by those same T cells as well as by natural killer cells [1]. GVHD, which occurs in both acute and chronic forms, remains the major cause of death without relapse of primary disease or nonrelapse mortality (NRM) [2,3]. The primary treatment of acute GVHD, high dose systemic corticosteroids, has not changed in 40 years and only a third of patients achieve durable responses to initial corticosteroid therapy; survival among the remaining patients is poor [4].

A major obstacle to any new primary treatment for GVHD is the inability to determine the risk of GVHD for an individual patient at the onset of disease symptoms. The risk of mortality from GVHD correlates with its maximum clinical severity, but maximal severity can only be assigned retrospectively, after the response to treatment is known [5–7]. Thus, at the onset of GVHD symptoms the risk is not yet known, and most patients are treated similarly with high-dose corticosteroids (at least 1 mg/kg/day). The result is that a substantial number of patients are either undertreated and their GVHD progresses, or overtreated because their disease would have responded to lower doses. Unfortunately, the

large majority (70–90%) of patients who develop steroid-resistant acute GVHD will die either from the disease or infections that occurred as a result of prolonged immunosuppression [8,9]. The toxicity from overtreatment is not limited to infection risks and includes other substantial morbidities such as avascular necrosis (sometimes requiring joint replacement) and diabetes mellitus.

A variety of biomarkers have been shown to predict acute GVHD severity with various degrees of accuracy. Unfortunately, most biomarkers have not been consistently successful or reproducible [10–14] with a noticeable lack of consistency of individual biomarkers across different centers. Until recently there were no studies validating the performance of a biomarker in a multicenter dataset, but last year our group published a scoring system that used an algorithm of three biomarkers (TNFR1, ST2 and REG3a) [15]. This scoring system consistently classified patients into three categories of risk that predicted response to treatment within 28 days as well as risk of mortality from acute GVHD in a training set (n = 328), and two validation sets (n = 164 and n = 300).

The algorithm that defined the risk was developed with the biomarker concentrations assessed for an individual patient at the onset of clinical symptoms of acute GVHD. Patient samples in the training set and the test set were taken from the biorepositories of two institutions (University of Michigan and



**Anne S Renteria**

The Tisch Cancer Institute – Bone Marrow & Stem Cell Transplantation, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA



**John E Levine**

The Tisch Cancer Institute – Bone Marrow & Stem Cell Transplantation, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA



**James L Ferrara**

Author for correspondence:  
Hematologic Malignancies Translational Research Center, The Tisch Cancer Institute & Division of Hematology/ Medical Oncology, Hess Center for Science & Medicine, New York, NY 10029, USA  
Tel.: +1 212 824 9365  
[james.ferrara@mssm.edu](mailto:james.ferrara@mssm.edu)

University of Regensburg, Germany) with different clinical practices regarding GVHD prophylaxis and BMT supportive care. The algorithm and the thresholds that demarcated risk categories were defined in the training set patients and then applied to the test set patients; similar incidences of NRM and rates of response to treatment within 28 days were observed.

“...the biomarker scoring provides information beyond the clinical risk factors alone and likely reflects target organ damage that is underway, even if not yet clinically manifest.”

The acid test of the algorithm was its validation in a third independent group of 300 patients (validation set) who had enrolled in one of two trials of primary treatment of acute GVHD conducted by the Blood and Marrow Transplant Clinical Trials Network (BMT CTN). These patients, from 18 different USA centers with a wide variety of BMT conditioning regimens, GVHD prophylaxis strategies and supportive care measures, had provided samples at the onset of therapy to a central BMT CTN biorepository. Once again, the scoring system successfully classified risk: patients with a low score (AA1) had only 8% mortality at 1 year, whereas AA2 patients had 27% mortality and AA3 patients had 46% mortality, with each group statistically significantly different from the next. Importantly, the cumulative incidence of relapse was not significantly different among the three groups, however the overall mortality differed significantly between each group. We interpreted this to mean that the biomarker concentrations did not reflect the strength of the GVH reaction but rather the actual damage to GVHD target tissue, especially in the gastrointestinal (GI) tract.

It is of great interest that the three biomarkers included in this algorithm possess biological relevance to GI GVHD. TNFR1, a surrogate for TNF- $\alpha$ , amplifies GI injury [16] and plasma levels increase before clinical manifestations of GVHD [17]. TNF- $\alpha$  regulates the expression ST2 that, together with its ligand IL-33, mediates inflammatory bowel disease [18]. The soluble form of ST2 is secreted in response to inflammatory stimuli, and IL-33 enhances inflammatory T cells that are involved in GVHD pathophysiology. ST2 levels in the plasma predict the lack of response to GVHD therapy and subsequent mortality [13]. REG3 $\alpha$ , an antimicrobial peptide produced by Paneth cells and other GI epithelia, and validated as a GI tract GVHD specific biomarker [12], protects GI epithelium from Gram-positive bacteria [19]. We postulate that the increased concentrations of these biomarkers in the blood at the time of GVHD onset reflect early damage to the GI tract that has not yet resulted in GI symptoms.

Several clinical risk factors can predict GVHD severity and survival in patients with GVHD [8,16,20]. AA1 predicted a lower risk of NRM than AA2, and AA3 higher risk than AA2 regardless of the presence of important pretransplant clinical risk factors such as donor type, age, conditioning regimen intensity or HLA-match. Indeed, patients with HLA-mismatched donors were significantly more likely to have AA3 GVHD in all three datasets. Thus the biomarker scoring provides information beyond the clinical risk factors alone and likely reflects target organ damage that is underway, even if not yet clinically manifest.

With regard to treatment decisions, GVHD grading systems cannot guide the choice of therapy at symptom onset, and therefore clinicians do not intensify immunosuppressive treatment of GVHD until primary therapy has failed. But the Ann Arbor algorithm defines three GVHD severity scores across the full range of clinical presentations and is therefore promising as a guide to treatment. Importantly, an elevated Ann Arbor score predicted later development of GI GVHD in patients who did not initially present with GI symptoms. Most deaths of patients with GVHD that are not caused by relapse of primary disease are due to poor response to treatment of GVHD in the GI tract. This observation supports the concept that this laboratory test can detect early subclinical disease and identify patients who need more intensive interventions.

This biomarker algorithm clearly advances previous work using single biomarkers but important limitations remain. First, although the algorithm predicts outcomes better than clinical symptoms, it still has fairly poor predictive power for the middle group of patients (AA2) whose overall risk has not changed; the algorithm's utility would be greatest for patients who score at either end (AA1 or AA3). Second, the algorithm's ability to guide prospective therapy is yet to be shown and must be tested in clinical trials. On such trials a low score might be used as an exclusion criterion for patients with severe clinical symptoms (e.g., voluminous diarrhea) from a trial of an investigational drug. In this case, patients who are likely to respond to standard therapy benefit by avoiding exposure to the risks of an experimental drug, and the trial also benefits by enrichment with patients who are less likely to respond to standard therapy. Alternatively, a low score could be used as an inclusion criterion to restrict exposure to lengthy corticosteroid regimens that are likely to result in complete response but that carry risks for opportunistic infections. Conversely, a high score (about 23% of all GVHD) could be used as an inclusion criterion for a trial of intensive primary therapy. This approach would be particularly beneficial for patients who present with mild symptoms but who are less likely to respond to standard therapy and who

might otherwise need to wait until primary treatment has failed before the initiation of a more intensive or experimental modality. Finally, a high score could serve as an exclusion criterion for a trial of low intensity therapy, which is very likely to fail in a patient with high risk disease. This approach could benefit both patients who are excluded (who may qualify for experimental, intensive therapy) and patients who are included (who are more likely to benefit from the reduction in therapy). Such an approach has been incorporated in a forthcoming trial of the BMT CTN. If a clinical trial is unavailable, a high score could lend confidence to the diagnosis of GVHD when a biopsy is equivocal, and a low score in a patient with a limited rash might support

the use of topical treatment or watchful waiting. We therefore expect this biomarker algorithm to become a powerful tool in the development of precision medicine for HCT patients.

#### Financial & competing interests disclosure

JL Ferrara and JE Levine have pending patents on biomarkers for diagnosis of graft-versus-host disease. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

#### References

- Ferrara JL, Levine JE, Reddy P, Holler E. Graft-versus-host disease. *Lancet* 373(9674), 1550–1561 (2009).
- Gooley TA, Chien JW, Pergam SA *et al.* Reduced mortality after allogeneic hematopoietic-cell transplantation. *N. Engl. J. Med.* 363(22), 2091–2101 (2010).
- Anasetti C, Logan BR, Lee SJ *et al.* Peripheral-blood stem cells versus bone marrow from unrelated donors. *N. Engl. J. Med.* 367(16), 1487–1496 (2012).
- Deeg HJ. How I treat refractory acute GVHD. *Blood* 109(10), 4119–4126 (2007).
- Rowlings PA, Przepiorka D, Klein JP *et al.* IBMTR Severity Index for grading acute graft-versus-host disease: retrospective comparison with Glucksberg grade. *Br. J. Haematol.* 97(4), 855–864 (1997).
- Cahn JY, Klein JP, Lee SJ *et al.* Prospective evaluation of 2 acute graft-versus-host (GVHD) grading systems: a joint Société Française de Greffe de Moëlle et Thérapie Cellulaire (SFGM-TC), Dana Farber Cancer Institute (DFCI), and International Bone Marrow Transplant Registry (IBMTR) pros. *Blood* 106(4), 1495–1500 (2005).
- Przepiorka D, Weisdorf D, Martin P *et al.* 1994 Consensus Conference on Acute GVHD Grading. *Bone Marrow Transplant.* 15(6), 825–828 (1995).
- Levine JE, Logan B, Wu J *et al.* Graft-versus-host disease treatment: predictors of survival. *Biol. Blood Marrow Transplant.* 16(12), 1693–1699 (2010).
- Westin JR, Saliba RM, De Lima M *et al.* Steroid-refractory acute GVHD: predictors and outcomes. *Adv. Hematol.* 2011, 601953 (2011).
- Paczesny S, Krijanovski OI, Braun TM *et al.* A biomarker panel for acute graft-versus-host disease. *Blood* 113(2), 273–278 (2009).
- Paczesny S, Braun TM, Levine JE *et al.* Elafin is a biomarker of graft-versus-host disease of the skin. *Sci. Transl. Med.* 2(13), 13ra2 (2010).
- Ferrara JL, Harris A, Greenson JK *et al.* Regenerating islet-derived 3- $\alpha$  is a biomarker of gastrointestinal graft-versus-host disease. *Blood* 118(25), 6702–6708 (2011).
- Vander Lugt M, Braun TM, Hanash S *et al.* ST2 as a marker for risk of therapy-resistant graft-versus-host disease and death. *N. Engl. J. Med.* 369(6), 529–539 (2013).
- Levine JE, Logan BR, Wu J *et al.* Acute graft-versus-host disease biomarkers measured during therapy can predict treatment outcomes: a Blood and Marrow Transplant Clinical Trials Network study. *Blood* 119(16), 3854–3860 (2012).
- Levine JE, Braun TM, Harris AC *et al.* A prognostic score for acute-graft-versus-host disease based on biomarkers: a multicenter study. *Lancet* 2(1), e21–e29 (2015).
- Schmaltz C, Alpdogan O, Muriglan SJ *et al.* Donor T cell-derived TNF is required for graft-versus-host disease and graft-versus-tumor activity after bone marrow transplantation. *Blood* 101(6), 2440–2445 (2003).
- Choi SW, Kitko CL, Braun T *et al.* Change in plasma tumor necrosis factor receptor 1 levels in the first week after myeloablative allogeneic transplantation correlates with severity and incidence of GVHD and survival. *Blood* 112(4), 1539–1542 (2008).
- Pastorelli L, Garg RR, Hoang SB *et al.* Epithelial-derived IL-33 and its receptor ST2 are dysregulated in ulcerative colitis and in experimental Th1/Th2 driven enteritis. *Proc. Natl Acad. Sci. USA* 107(17), 8017–8022 (2010).
- Ogawa H, Fukushima K, Naito H *et al.* Increased expression of HIP/PAP and regenerating gene III in human inflammatory bowel disease and a murine bacterial reconstitution model. *Inflamm. Bowel Dis.* 9(3), 162–170 (2003).
- Hurley CK, Woolfrey A, Wang T *et al.* The impact of HLA unidirectional mismatches on the outcome of myeloablative hematopoietic stem cell transplantation with unrelated donors. *Blood* 121(23), 4800–4806 (2013).